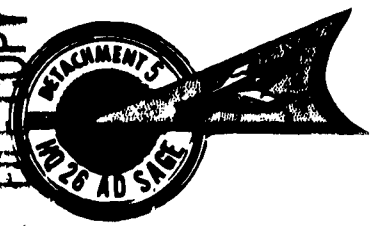

THE DEVELOPMENT OF JOB PROFICIENCY TESTS FOR SELECTED OPERATOR POSITIONS IN THE WEAPONS BRANCH

JOINT TEST STAFF

Category III Evaluation

P & T REPORT 27

(5) 741 500

(6) THE DEVELOPMENT OF JOB PROFICIENCY TESTS FOR SELECTED OPERATOR POSITIONS IN THE WEAPONS BRANCH

(9 thru 12) NH

(13) 21

(14) NP

REPORT NUMBER

(14) PRT-27

(11) JUNE 1960

(AFCCDD-TN-60-27)
(~~PRA Report~~ 60-32)

This is a Joint Test Staff, Detachment 5, 26th Air Division (SAGE) Personnel and Training Report accomplished by the Psychological Research Associates, Inc., under (contract AF19(604)-5616, monitored by OAO, AFCCDD and prepared in partial fulfillment of the requirements of SAGE Category III Evaluation (AFR 80-14).

(15)

James J. Jarman for
ROBERT W. WALTZ Col USAF
Colonel, USAF
Commander

(16 thru 19) NH

(20) 71

(21) NH

LIST OF TABLES

TABLE OF CONTENTS

INTRODUCTION

The research reported ~~herein~~ was concerned with the development of paper-and-pencil job knowledge tests for six positions in the Weapons Branch of a Direction Center in the SAGE System. ~~Actually,~~ three tests were developed, the positions being paired as follows: Senior Director/Senior Director Technician (SD/SDT); Weapons Director/Weapons Director Technician (WD/WDT); Intercept Director/Intercept Director Technician (IND/INT).

The report treats the job description techniques, rationale for test outline, item development, preliminary tryout in the New York Air Defense Sector (NYADS), item analysis and test revision, final administration at the Boston and Syracuse Air Defense Sectors (BOADS and SYADS), and at the training facility at Richards-Gebaur Air Force Base (RG). Also included are information on test reliability and validity, and recommendations for normative use of the tests.

The test materials, including the items, and instructions for administering, scoring, and interpreting the results, are printed in a separate booklet.

TEST DEVELOPMENT

The test development phase of the contract was divided into four sub-phases: job description, item-writing, trial administration and test revision, and test analysis. These sub-phases are considered in detail below.

JOB DESCRIPTION

The technique for describing the positions prior to the construction of test items was that of the task-equipment analysis (TEA). As the term implies, the analysis describes the relationship between the equipment to be operated and the task of the operator. A detailed description of the philosophy and methodology of the TEA may be found in AFCRC-TN-59-76, "SAGE Task Equipment Analysis—Intercept Director/Intercept Director Technician," February 1960, or in any of the TEAs covering other positions in the SAGE Direction Center.[1]

ITEM-WRITING

Another contractor has developed paper and pencil tests for all Direction Center Operator Positions other than those in the Weapons Branch.[2] An examination of the TEAs for the SASO and SC in the Combat Center revealed that these jobs were unsuitable as subjects for paper-and-pencil test development. There were several reasons why this is so. First, there was very little job content in day-to-day operation, the tasks being mostly limited to decision-making at the time of exercises or missions. Second, the number of incumbents was insufficient for any psychometric analysis; hence, there would

[1]Other Technical Notes published as TNs under the auspices of the Operational Applications Laboratory, Air Force Cambridge Research Center or the Operational Applications Office, Air Force Command and Control Development Division.

[2]See AFCRC-TN-58-63, 58-64, 58-65, 58-66, 58-62.

be no way of checking on the characteristics of the tests produced. For these reasons it was determined, with the consent of OAL, that test development would be limited to six positions in the Weapons Branch: Senior Director (SD) and Senior Director Technician (SDT), Weapons Director (WD) and Weapons Director Technician (WD), and Intercept Director (IND) and Intercept Director Technician(INT).

The original test plan involved the development of five types of test items:

1. S items, involving knowledge of Situational Displays and Digital Display symbology
2. C items, involving knowledge of computer capabilities and functions
3. T items, involving knowledge of tactics, SOPs and aircraft and weapons capabilities
4. D_1 items, involving tactical decisions
5. D_2 items, involving decisions affecting areas broader than tactics

It soon became evident that D_1 and D_2 item types were not readily amenable to paper-and-pencil, multiple choice format. Possible items of these types were too few to constitute a realistic part of any test. Accordingly, it was determined to limit the items to S, C, and T types.

For the S items, the use of photographic reproductions of SIDs and DIDs was considered and rejected. Observation of the operators indicated that they were able to read symbology very easily, despite the fact that the dynamic quality of the displays seems to pose a serious problem. Thus, photographic reproduction might add a simulation of reality unrelated to the actual task, i.e., the interpretation of the symbology.

Using the TEA and other sources, approximately 350 multiple choice items were written. Many of these were, of course, appropriate to several positions in the Weapons Branch. The actual distributions of items by type among the positions are given in Table 1, Each of the items developed was reviewed by a military subject matter expert; an additional review was performed for psychometric propriety.

An attempt was made to anticipate situations which would make items inapplicable or obsolete. A scheme was devised whereby each item was classified according to its sector applicability, and susceptibility to error because of program model, weapons, or tactics changes. Whenever such changes occurred or the test was to be used in a new sector, it was planned that the indicated items would be reviewed. It became apparent that it was not possible to apply such classifications reliably. If such a system were used, it would be quite possible that some "bad" items would be overlooked. Alternatively, the wisest policy seemed to be that of reviewing the entire test by a subject matter expert before it is applied in a particular sector situation.

TRIAL ADMINISTRATION AND TEST REVISION

The six tests, with item composition as shown in Table 1, were administered at NYADS, McGuire Air Force Base, New Jersey. The results of that testing indicating means and standard deviations for each position, are given in Table 2. The means are sufficiently low in relation to maximum possible scores to be discriminating; the standard deviations are consistent with the small Ns.

It may be considered as a maxim that item analyses for achievement test items require an N of at least 200, and that computation of validity and reliability coefficients requires an N of at least 75, for the results to be meaningfully interpretable. Nevertheless, because it was felt that some sort of psychometric information was necessary before proceeding further in test development, such analyses were performed on the data for INDs (N = 19) and INTs (N = 15). It was

TABLE 1

NUMBER OF ITEMS BY TYPE FOR OPERATOR POSITION PROFICIENCY TESTS IN THE SAGE SYSTEM

OPERATOR POSITION	ITEM TYPE			TOTAL
	S	C	T	
SD	51	56	48	155
SDT	50	33	33	116
WD	48	80	77	205
WDT	50	47	50	147
IND	46	70	82	198
INT	42	34	47	133

TABLE 2

SUMMARY OF RESULTS OF PROFICIENCY TEST ADMINISTRATION IN NYADS

OPERATOR POSITION	N	$\bar{X}$	s	MAX.
SD	4	124.3	2.6	155
SDT	5	93.6	2.8	116
WD	2	—	—	—
WDT	7	113.4	7.6	197
IND	20	149.6	11.0	198
INT	15	94.1	7.2	133

possible to get ranking information from the superiors of 15 of the INDs. Inasmuch as no ranker was aware of performance in all three weapons teams, this necessitated triple ties at each rank, since each Weapons Director ranked the INDs in his team from 1 to 5. These rankings were used as criteria in a validity check of the total test and each of its parts. It was not possible to get estimates of the reliability of these rankings, since only one ranker was available for each group of INDs. The results of the test reliability and validity analyses are given in Table 3. Although the test reliability estimates using Kuder-Richardson Formula 20 were quite low in some instances, the test-part intercorrelations and external validity coefficient for the INDs were deemed to be satisfactory evidence of the adequacy of the test items.

The validity index (phi coefficient) and the p-value of each item were examined for the IND and INT data. All items with negative phi coefficients were discarded. Also, the limits 0.15 and 0.85 were established arbitrarily, and items with p-values below or above these limits, respectively, were discarded. This procedure was adopted since items with p-values more extreme than these limits contribute little to item reliability and test validity. No analysis was possible for the data from the SD/SDT and WD/WDT test administrations since too few cases were available.

A re-examination of the mission of the paper-and-pencil proficiency tests led to the conclusion that there was no need to have separate tests for the operator and technician at any one position. Essentially, each is supposed to know all of the information (as distinguished from skills and decision-making ability) required for the combined position. It might be expected, of course, that the technicians would, on the average, earn lower scores than the operators when both are administered the same test. This expectation seems reasonable since operators are commissioned officers and technicians are airmen, and the former group is both more directly involved in the job and has the advantage

TABLE 3

RESULTS OF RELIABILITY AND VALIDITY ANALYSES OF PROFICIENCY TESTS ADMINISTERED TO INDs AND INTs IN NYADS

Operator Position	Item Type	$\bar{X}$	s	KR_{20}	Max. Score	C	r T	Criteron
IND	S	11.73	1.96	.15	17	.33	.20	
IND	C	24.10	3.51	.37	41		.51	
IND	T	26.16	3.65	.29	47			.49
TOTAL		64.6	6.8	.55	105			.61
INT	S	12.7	1.98	.05	19	.39	.48	
INT	C	17.0	3.81	.63	30		.34	
INT	T	21.1	4.38	.62	39			
TOTAL		50.9	7.7	.72	88			

of a better educational background. It seemed to be true also that some of the items were appropriate to all three positions in the Weapons Branch (SD/SDT, WD/WDT, and IND/INT), some to two of these positions, and some to only one position. Accordingly, the available items were categorized into six parts. Part I consisted of items common to all three positions, Part II for WD/WDT and IND/INT, Part III for IND/INT only, Part IV for SD/SDT and WD/WDT, Part V for WD/WDT only, and Part VI for SD/SDT only. The summary of the parts applicable to each position appears in Table 4.

TABLE 4

COMPOSITION OF SAGE PROFICIENCY TESTS BY MODULAR PARTS FOR THREE OPERATOR POSITIONS

SD/SDT		WD/WDT		IND/INT	
Form A	Form B	Form A	Form B	Form A	Form B
IA	IB	IA	IB	IA	IB
IVA	IVB	IIA	IIB	IIA	IIB
VIA	VIB	IVA	IVB	IIIA	IIIB
		VA	VB		

It is important to remember that these part numbers do not correspond in any way to the symbology, computer and tactic breakdown of the subject matter of the items. The part numbers merely represent modules which can be combined in various ways to produce tests for various positions. The combination of these parts into tests, giving the number of items in each part, is represented in Table 5. Since alternate forms were required, and the number of reliable, valid items was limited, it was decided to make some items common to both of the alternate forms required. Thus, for example, Table 5 indicates that Part I consists of 53 items for either form; however, in each form, 38 items are unique and 15 items are common.

TABLE 5

DISTRIBUTION OF ITEMS BY PARTS
FOR ALTERNATE OPERATOR TEST FORMS

OPERATOR POSITION TEST	PARTS	NO. ITEMS	COMMON	UNIQUE
SD/SDT	I	53	15	38
	IV	21	11	10
	VI	6	3	3
TOTAL		80	29	51
WD/WDT	I	53	15	38
	II	17	11	6
	IV	21	11	10
	V	10	4	6
TOTAL		101	41	60
IND/INT	I	53	15	38
	II	17	11	6
	III	12	6	6
TOTAL		82	32	50

TEST ANALYSIS

Administration

The revised tests were administered in the Weapons Branch at BOADS and SYADS and to graduating students of classes in Intercept Direction and Weapons Direction at Richards-Gebaur Air Force Base. The N's, means[1] and standard deviations for each part and the total test are reported in Tables 6 and 7. "Parts" as used here means the three subject matter categories of the test, i.e., symbology (S), computer knowledge (C), and tactics (T). At BOADS almost all incumbents took both the A and B forms. The testing order was counterbalanced so that A and B forms were given first and second equally often. At SYADS and Richards-Gebaur AFB, it was possible to test each incumbent with only one form; at these sites half of the incumbents were assigned each of the forms at random.

Analysis of Variance of Means

An analysis of variance was performed on the variation of means by operators, sectors and forms. Results are reported in Table 8. None of the main effects or interactions were found to be significant at the 5% level of confidence. This finding may be interpreted to mean that, for the number of test scores available, sector differences, operator differences and form differences observable in Table 6 may be attributed to random variation around an over-all average.

Reliability Analyses

A series of reliability analyses were performed. These were of several types. First, internal consistency analyses using the Kuder-Richardson Formula No. 20, or its equivalent the Hoyt analysis of

[1]Some of the means of Table 6 have been adjusted to reflect revision in two items which were scored erroneously.

TABLE 6

ADJUSTED MEAN SCORES ON SD/SDT, WD/WDT AND IND/INT PROFICIENCY TESTS BY PARTS AND ALTERNATE FORMS AT BOADS, RICHARDS-GEBAUR AFB AND SYADS

		BOADS					RICHARDS-GEBAUR					SYADS				
OPTR POSIT	TEST FORM		TEST PART					TEST PART					TEST PART			
		N	S	C	T	TOTAL	N	S	C	T	TOTAL	N	S	C	T	TOTAL
SD	A	5	16.80	17.60	21.80	61.60						2	12.50	14.00	18.50	49.50
SDT	A	5	16.00	16.20	21.80	59.20						3	17.00	13.66	15.00	49.33
SD	B	5	16.57	17.52	23.00	61.99						5	12.97	13.80	17.80	49.57
SDT	B	5	16.17	16.52	22.40	60.17						5	13.57	14.20	13.80	46.87
WD	A	6	16.66	20.17	24.51	78.17	11	16.26	25.17	20.55	70.54	2	16.50	23.01	23.01	69.00
WDT	A	4	18.00	25.74	22.26	74.50	9	15.88	21.78	17.43	60.28	2	16.50	23.01	20.01	67.50
WD	B	5	16.97	26.91	23.61	73.69	12	17.37	24.63	19.35	69.52	4	15.47	26.46	25.50	69.19
WDT	B	4	16.71	23.97	23.76	72.19	10	17.07	20.91	17.79	62.99	2	12.97	19.71	18.00	57.69
IND	A	14	10.21	24.63	23.22	58.07	17	10.18	22.41	23.28	55.88	12	9.67	22.08	19.17	51.00
INT	A	22	10.36	22.86	18.63	51.86	15	10.27	19.20	17.46	46.73	12	9.25	18.90	14.85	43.00
IND	B	14	10.97	25.08	25.44	61.12	14	10.90	20.43	21.12	52.48	14	8.40	20.91	21.36	50.69
INT	B	21	9.87	22.29	18.51	50.69	15	9.77	19.38	15.72	44.89	11	8.42	21.54	17.01	46.97

TABLE 7

STANDARD DEVIATIONS OF SCORES ON SD/SDT, WD/WDT AND IND/INT PROFICIENCY TEST BY PARTS AND ALTERNATE FORMS AT BOADS, RICHARDS-GEBAUR AND SYADS

		BOADS					RICHARDS-GEBAUR					SYADS					TOTAL
OPTR POSIT	TEST FORM		TEST PART					TEST PART					TEST PART				
		N	S	C	T	TOTAL	N	S	C	T	TOTAL	N	S	C	T	TOTAL	
SD	A	5	1.33	2.58	3.87	7.39						2	3.50	1.00	0.50	3.50	8.51
SDT	A	5	1.79	2.56	1.72	4.71						3	0.00	2.62	1.63	2.05	6.19
SD	B	5	1.36	3.37	2.28	5.93						5	4.04	3.49	3.37	9.89	10.23
SDT	B	5	1.94	2.71	3.83	8.30						5	2.24	2.31	2.97	5.04	9.59
WD	A	6	1.49	2.79	1.50	6.12	11	1.29	3.16	2.43	4.29	2	1.50	1.00	2.00	2.00	6.04
WDT	A	4	0.71	1.79	3.11	4.72	9	1.72	3.55	3.17	7.54	2	0.50	3.00	1.00	4.50	9.00
WD	B	5	1.41	2.56	1.02	4.98	12	0.76	1.93	4.73	5.65	4	1.66	1.92	1.12	2.49	5.83
WDT	B	4	1.09	3.42	4.32	8.26	10	1.58	3.40	2.68	6.89	2	1.00	1.00	3.00	6.00	8.56
IND	A	14	1.47	2.84	3.17	4.59	17	0.98	2.33	3.18	4.44	12	1.49	3.12	3.26	6.90	5.99
INT	A	22	0.83	2.99	3.26	5.93	15	1.18	4.45	2.52	5.01	12	1.16	2.96	3.83	6.87	6.50
IND	B	14	1.00	2.16	3.48	3.35	14	1.71	2.40	3.18	4.21	14	2.13	3.23	3.60	7.40	5.50
INT	B	21	1.57	2.87	4.17	6.14	15	1.42	3.20	3.77	6.18	11	2.02	3.95	4.11	8.66	7.29

TABLE 8

ANALYSIS OF VARIANCE OF IND/INT
PROFICIENCY TEST SCORES
SECTORS X OPERATORS X FORMS

SOURCE	SUM OF SQUARES	d.f.	MEAN SQUARE	F
Sectors	1362	2	681	
Operators	2174	1	2174	3.67
Forms	173	1	173	
S x O	536	2	268	
O x F	9	1	9	
S x F	191	2	95	
S x O x F	817	2	408	
Within	102768	169	608	
	104321	176	593	

variance technique.[1] These were computed for the total test, each position and each form separately, by sites. Using the same technique, the reliabilities were also computed by pooling the operator and technician scores at each position, and then pooling these scores by sectors in order to increase the N. Results for the internal consistency reliability computations are presented in Table 9; results for the second two analyses are shown in Tables 10 and 11. The small number of cases associated with each coefficient should be considered in any interpretation, since in some instances negative coefficients and a coefficient greater than unity occured. The coefficient of −1.07 is correctly computed. It is possible in either of these methods of reliability estimation to obtain much anomalous results. This can be

[1]Hoyt, C. Test reliability obtained by analysis of variance. Psychometrika, 1941, 6, 153-160.

TABLE 9

RELIABILITY COEFFICIENTS COMPUTED BY KUDER RICHARDSON FORMULA 20 FOR PROFICIENCY TESTS ADMINISTERED IN ALTERNATE FORMS AT THREE SITES

OPERATOR POSITION	SITE											
	BOADS				RICHARDS GEBAUR				SYADS			
	Form		Form		Form		Form		Form		Form	
	N	A	N	B	N	A	N	B	N	A	N	B
SD	5	.87	5	.77					2	.41	5	.86
SDT	5	.63	5	.86					3	-1.07	5	.42
WD	6	.71	5	.55	11	.24	12	.58	2	-.51	4	-.57
WDT	4	.52	4	.83	9	.72	10	.69	2	.65	2	.72
IND	14	.42	14	.10	17	.42	14	.35	12	.76	14	.74
INT	22	.62	21	.63	15	.44	15	.61	12	.75	11	.81

TABLE 10

RELIABILITY COEFFICIENTS COMPUTED BY KUDER RICHARDSON FORMULA 20 FOR POOLED DATA ON PROFICIENCY TESTING AT THREE SITES

DATA POOL	FORM	R_{tt}	N
BOADS and SYADS	A	.80	15
SD and SDT	B	.87	20
BOADS, SYADS and R-G	A	.77	34
WD and WDT	B	.74	37
BOADS, SYADS and R-G	A	.74	92
IND and INT	B	.76	89

TABLE 11

ALTERNATE FORM RELIABILITY COEFFICIENTS FOR PROFICIENCY TESTING AT THREE POSITIONS IN BOADS

OPERATOR TEST	N	r_{AB}
SD	5	.68
SDT	5	.58
SD and SDT	10	.60
WD	5	.90
WDT	4	.91
WD and WDT	9	.81
IND	14	.60
INT	20	.40
IND and INT	34	.64

exemplified in the Hoyt approach where it is only necessary that the error variance exceed the variance among individuals. It should be remembered also that this coefficient is based on an N of 3; the rational explanation is that it is as though these 3 individuals took a test twice and reversed the rank order of their scores which ranged from 47 to 52. Second, alternate form reliabilities were computed for the BOADS data, since in that sector each incumbent took both the A and B forms of the test. For 27 cases where test order was AB, the alternate form reliability coefficient was 0.60; for 26 cases in which the order was BA, the coefficient was 0.44. Third, for those items which were common to both forms, a reliability coefficient was computed for the test-retest results using pooled IND, INT scores from BOADS. This coefficient was 0.59. This is interesting since it is not appreciably better than those obtained within alternate forms when only some of the items were common to both forms. Finally, using the INT data from BOADS only, the reliabilities of the S, C and T parts of the test were estimated using the analysis of variance technique of Hoyt. This procedure is identical in result to K-R Formula 20. The obtained reliabilities were 0.18(S), 0.43 (C), and 0.46 (T). There is some question, however, about the satisfaction of the assumptions required for making this type of reliability estimate with either the KR-20 or Hoyt technique. One such assumption is that of complete homogeneity of item content within parts; stated in another way, that there is no variance common to groups of items within a part. Because of this reservation the figures given should be interpreted as lower bounds. Unfortunately, no upper bound can be estimated.

Part Intercorrelation

Intercorrelations were computed among the S, C, and T parts of the test. Only the IND and INT data were used because too few cases were available for the other positions. For BOADS the coefficients represent pooled results for both forms of the test. The results are reported in Table 12. The NYADS data are from the pilot administration

TABLE 12

INTERCORRELATIONS OF PARTS OF SAGE PROFICIENCY TESTS ADMINISTERED TO IND's AND INT's AT DIFFERENT SITES

SITES	IND			INT			IND/INT		
	N = 44			N = 28			N = 72		
BOADS		C	T		C	T		C	T
	S	43	26	S	03	05	S	35	26
	C		29	C		-21	C		22
	N = 30			N = 30			N = 60		
RICHARDS-GEBAUR		C	T		C	T		C	T
	S	08	18	S	-07	27	S	08	26
	C		39	C		17	C		39
	N = 23			N = 26			N = 49		
SYADS		C	T		C	T		C	T
	S	27	41	S	60	31	S	44	31
	C		63	C		41	C		54
	N = 15			N = 19					
NYADS*		C	T		C	T			
	S	39	48	S	33	20			
	C		34	C		51			
	N = 97			N = 84			N = 101		
ALL**		C	T		C	T		C	T
	S	29	33	S	31	33	S	31	32
	C		48	C		29	C		45

* Data from preliminary longer form of test

** Does not include NYADS data because the method of reliability estimation does not permit pooling results from tests with differing numbers of items.

with a longer test, before item analysis. If the coefficients for the combined data based on 181 cases are taken as the best estimates of the parameter values, then only one of the 39 other coefficients is different to a degree significant at the 5% level of confidence. Thus any interpretations of differences existing between operators or among sectors is unwarranted.

Validity Analyses

At BOADS and SYADS it was possible to obtain a supervisor's ranking for both IND's and INT's participating. At SYADS the ranking was accomplished within crews; the resulting forced ties presumably attenuated correlations between these and other data. The validity coefficients for parts and test total are reported in Table 13. Once again, the very small number of cases precludes adequate interpretation. Even with the negative coefficients obtained in some cases, differences are not significant at the 5% level of confidence for any pair of part validity coefficients, or any pair of total validity coefficients.

Additional Items

For the test as described, all items dealing with aircraft pertain to the F-106A. To meet the needs for other aircraft, depending on the sector in which the test is administered, 10 items each of parallel form have been prepared for the F-86L, F-89J, F-101B, F-102A, and F-104A. Since these items have not been administered, their psychometric characteristics are unknown. However, it is believed that, aside from fluctuation in p-value associated with the relative accessibility of this kind of information in various sectors, these items are essentially the same as those used in Forms A and B of the test.

Item p-values

The percentage of examinees succeeding on each item was computed for all items in the IND/INT test, for IND's and INT's separately at

each of the sites. These data are given in the Appendix. Extreme caution should be used in interpreting these p-values, since the N's range from 11 to 22. It will be noted that, despite the original item analysis, 14 of the items in Form A and 11 of the items in Form B have p-values greater than 0.85.

TABLE 13

VALIDITY COEFFICIENTS USING A RANKING CRITERION FOR A PROFICIENCY TEST ADMINISTERED AT BOADS AND SYADS TO IND's AND INT's

			Test Part			
SITE	POSITION	N	S	C	T	TOTAL
BOADS	IND	14	.21	-.22	.31	.28
SYADS	IND	25	.45	.22	-.03	.34
BOADS	INT	13	.12	.40	.13	.26
SYADS	INT	23	.37	.33	.32	.44

DISCUSSION

From the information presented, it is possible to conclude that the tests provide discrimination among individuals, that they are reliable, that they have satisfactory validity against an external criterion, and that the alternate forms are reasonably equivalent. It is unfortunate, however, that it is not possible to be definitive in the interpretation of differences occurring among the data. The differences, between operator and technician in the same position, between different sectors for the same position, and between the alternate forms, were not statistically significant, using the IND/INT data. If the failure to reach significance is a function of the size of the samples, this will never be known since, in each instance, the samples include all of the available incumbents.

In the opinion of the writer, real differences do exist among the data, and it is believed that these would become evident if there were larger samples. For example, in every one of the 6 comparisons possible between an IND and an INT on total score, the operator had the higher score. Since, if operator and technician test performances were really equal, the cited event could be expected to occur by chance on 1 in 64 times, it is reasonable to conclude that the operators have more test knowledge by some slight amount. It is true also that for the SD vs SDT and WD vs WDT comparisons, the operator has a higher score in each case. This event could be expected by chance only 1 in 2,048 times, so again it is reasonable to conclude that operators have the greater amount of test knowledge.

Similarly, in comparing sectors, in every one of the 36 possible comparisons on part scores, and the 12 possible comparisons on total scores, the BOADS incumbents scored higher than those in SYADS. In comparing BOADS and Richards-Gebaur incumbents the BOADS personnel have higher scores on part-scores in 23 of 24 instances, and higher total scores in every one of 8 instances. In contrast, for Richards-Gebaur vs SYADS, the Richards-Gebaur personnel are higher in 14 of

24 part-score comparisons, and in 6 of 8 total score comparisons. These results are consistent with chance fluctuation from a true equal score value, although the RG sample consists of students and the SYADS sample is operating personnel. It seems reasonable to conclude, therefore, that SYADS and Richards-Gebaur personnel have less of the knowledge measured by the test than do the personnel at BOADS.

It is suggested on the basis of the foregoing discussion that the data on test means from the BOADS testing together with the variability data based on the pooled results from all three sites be used for normative purposes. A lower bound of acceptability might be defined as 2 standard deviations below the mean for any particular position. Such a score would compare approximately with a stanine score of 1. For example, IND's might be required to earn a total score of 48. Using the average standard deviation for both forms of IND test (5.75), the computation would be 59.59 (the mean for both forms of the IND test) minus 11.50 (2 x 5.75) = 48.09. Application of the cut-off techniques to the S, C, and T part-scores of the test is not recommended at this time. Although the part reliabilities of 0.18, 0.43 and 0.46 are lower bounds, there is no evidence in the absence of further data collection that part-scores are sufficiently reliable for individual test interpretation.

APPENDIX

PERCENTAGE OF IND/INT EXAMINEES SUCCEEDING ON EACH ITEM OF THE FORM A TEST AT EACH OF THREE SITES

PART	BOADS		R-G		SYADS		
	N 14	N 22	N 17	N 15	N 12	N 12	
	IND	INT	IND	INT	IND	INT	$\bar{p}$
I- 1	79	86	94	93	100	67	87
2	71	91	100	87	92	33	82
3	93	100	100	100	100	92	98
4	93	100	94	100	100	100	98
5	43	41	18	53	25	67	40
6	86	82	59	71	67	83	75
7	86	82	76	82	75	67	79
8	100	95	88	76	100	100	93
9	100	95	94	93	100	83	97
10	100	95	100	100	100	100	99
11	93	77	88	67	50	64	75
12	71	77	94	67	58	50	72
13	14	32	0	53	25	42	27
14	93	77	47	47	92	83	72
15	93	95	100	100	100	83	96
16	79	73	100	60	92	67	79
17	71	41	100	40	58	33	68
18	93	86	100	73	100	92	90
19	50	59	88	100	67	83	74
20	71	82	76	100	92	75	83
21	64	59	76	47	33	08	51
22	64	64	71	73	50	42	62
23	57	32	100	20	25	17	44
24	50	55	76	20	58	33	50

PART	BOADS N 14 IND	BOADS N 22 INT	R-G N 17 IND	R-G N 15 INT	SYADS N 12 IND	SYADS N 12 INT	$\bar{p}$
25	50	41	47	60	08	08	38
26	64	82	06	27	100	100	61
27	36	36	29	40	17	08	29
28	43	27	47	27	25	33	34
29	79	68	76	40	33	25	56
30	43	41	82	27	92	33	52
31	36	41	88	53	75	75	60
32	93	50	76	33	42	33	55
33	57	64	71	53	75	58	63
34	64	82	82	73	58	75	74
35	100	91	100	100	100	100	98
36	64	64	65	53	75	67	64
37	57	27	41	13	50	33	34
38	71	59	47	40	25	08	44
39	79	50	73	47	83	42	61
40	0	09	12	33	0	0	10
41	93	86	88	87	67	67	83
42	86	91	100	80	0	33	71
43	86	91	73	80	83	42	81
44	93	82	100	80	75	83	86
45	93	77	24	27	50	08	49
46	50	36	53	27	50	25	38
47	43	27	59	20	25	17	33
48	57	32	71	47	42	33	47
49	86	55	65	73	100	92	75
50	21	14	47	33	25	25	27
51	64	59	65	27	75	83	61
52	79	50	76	60	67	58	64

PART	BOADS N 14 IND	BOADS N 22 INT	R-G N 17 IND	R-G N 15 INT	SYADS N 12 IND	SYADS N 12 INT	$\bar{p}$
53	0	09	18	13	08	17	9
II- 1	100	91	59	67	100	83	83
2	100	100	88	47	100	100	89
3	100	91	35	67	92	100	79
4	57	45	41	60	50	100	56
5	86	68	53	53	67	25	60
6	93	82	76	40	58	17	64
7	86	68	47	60	67	42	62
8	64	55	35	67	33	42	50
9	100	91	100	82	92	58	88
10	79	77	88	53	67	92	76
11	86	86	82	80	58	17	72
12	57	91	100	100	100	100	91
13	50	41	53	33	42	50	45
14	36	18	12	20	17	17	20
15	93	23	65	13	75	33	48
16	71	27	47	40	42	25	41
17	93	45	71	53	75	50	63
III- 1	100	91	82	82	75	92	87
2	100	100	71	20	100	92	81
3	71	82	76	80	75	83	78
4	71	82	88	60	75	67	75
5	93	100	100	60	83	75	87
6	57	23	65	27	25	08	35
7	79	41	47	60	25	08	45
8	79	73	88	73	92	83	81
9	36	36	35	07	17	25	27
10	71	55	76	73	50	25	60
11	57	27	41	40	33	25	37
12	43	45	35	27	58	50	42

PERCENTAGE OF IND/INT EXAMINEES SUCCEEDING ON EACH ITEM OF THE FORM B TEST AT EACH OF THREE SITES

PART	BOADS		R-G		SYADS		
	N 14	N 21	N 14	N 15	N 14	N 11	
	IND	INT	IND	INT	IND	INT	$\bar{p}$
I—1	79	71	100	67	57	64	73
2	93	81	100	80	50	45	78
3	79	57	86	73	64	82	72
4	100	95	93	100	50	55	84
5	86	86	100	73	71	73	82
6	100	100	100	100	93	100	99
7	93	57	79	87	57	27	67
8	100	100	100	80	93	91	94
9	100	95	94	93	100	83	97
10	86	71	93	87	50	64	75
11	100	100	100	93	86	100	97
12	86	71	71	47	71	27	64
13	14	19	07	60	50	64	34
14	100	86	64	40	71	91	75
15	79	62	71	67	86	64	71
16	79	57	79	80	71	73	72
17	79	100	71	73	64	91	81
18	100	100	79	80	93	82	90
19	86	57	93	40	93	64	70
20	100	86	93	100	57	82	87
21	14	43	14	40	07	55	29
22	43	38	43	73	43	55	48
23	100	33	100	29	71	55	62
24	79	57	57	20	29	55	50
25	43	29	29	29	21	27	30
26	79	86	14	14	64	45	35

	BOADS		R-G		SYADS		
	N 14	N 21	N 14	N 15	N 14	N 11	
PART	IND	INT	IND	INT	IND	INT	$\bar{p}$
27	43	57	14	67	14	18	31
28	43	48	79	33	57	55	52
29	71	33	71	14	64	09	44
30	86	48	57	40	86	36	59
31	71	29	29	21	36	36	36
32	100	76	100	60	100	90	86
33	79	71	71	67	86	64	73
34	79	43	57	80	64	27	58
35	93	76	21	14	55	27	50
36	100	67	71	53	71	55	70
37	50	38	71	33	55	27	44
38	79	67	64	33	71	55	62
39	71	82	64	80	79	82	72
40	93	52	57	40	86	73	65
41	100	95	100	73	100	82	92
42	100	86	79	67	100	100	88
43	100	67	86	29	00	09	51
44	93	38	86	53	93	73	70
45	29	38	43	40	57	55	43
46	36	38	50	40	29	45	39
47	29	29	50	47	21	09	35
48	71	33	43	29	57	51	46
49	64	43	93	67	36	45	57
50	29	29	43	43	29	18	32
51	64	57	73	33	50	27	52
52	71	52	79	33	71	27	56
53	09	14	07	00	14	00	08

	BOADS		R-G		SYADS		
	N 14	N 21	N 14	N 15	N 14	N 11	
PART	IND	INT	IND	INT	IND	INT	$\bar{p}$
II—1	93	100	50	67	79	64	78
2	93	100	86	80	93	91	91
3	100	100	73	73	64	73	82
4	93	52	86	47	86	73	71
5	57	57	14	33	64	64	48
6	14	29	00	53	43	45	30
7	64	67	43	67	43	64	59
8	93	67	79	67	71	73	74
9	93	81	100	73	93	91	88
10	93	81	86	87	86	91	87
11	57	91	100	100	100	100	91
12	79	67	36	87	64	73	68
13	57	47	73	33	64	45	53
14	50	24	07	20	36	18	26
15	100	76	100	43	100	64	80
16	71	48	36	20	36	18	39
17	93	67	73	53	71	55	69
III—1	100	90	100	87	93	82	92
2	100	76	93	47	71	55	74
3	93	90	91	60	64	64	78
4	71	62	79	67	64	82	59
5	93	86	91	67	86	82	84
6	57	33	91	20	29	27	42
7	79	71	64	40	14	36	53
8	86	86	91	80	79	91	85
9	50	24	07	13	07	55	25
10	50	52	64	60	29	18	47
11	64	33	21	47	36	27	38
12	14	24	14	29	50	73	32